

Mixing and matching: using qualitative methods to improve quantitative impact evaluations (IEs) and systematic reviews (SRs) of development outcomes

Jimenez, Emmanuel; Waddington, Hugh; Goel, Neeta; Prost, Audrey ; Pullin, Andrew; White, Howard; Lahiri, Shaon; Narain, Anmol

Journal of Development Effectiveness

DOI:

[10.1080/19439342.2018.1534875](https://doi.org/10.1080/19439342.2018.1534875)

Published: 30/11/2018

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](https://doi.org/10.1080/19439342.2018.1534875)

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):

Jimenez, E., Waddington, H., Goel, N., Prost, A., Pullin, A., White, H., Lahiri, S., & Narain, A. (2018). Mixing and matching: using qualitative methods to improve quantitative impact evaluations (IEs) and systematic reviews (SRs) of development outcomes. *Journal of Development Effectiveness*, 10(4), 400-421. <https://doi.org/10.1080/19439342.2018.1534875>

Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Mixing and matching: using qualitative methods to improve quantitative impact evaluations (IEs) and systematic reviews (SRs) of development outcomes

Emmanuel Jimenez, Hugh Waddington, Neeta Goel, Audrey Prost, Andrew Pullin, Howard White, Shaon Lahiri & Anmol Narain with the assistance of Raag Bhatia

To cite this article: Emmanuel Jimenez, Hugh Waddington, Neeta Goel, Audrey Prost, Andrew Pullin, Howard White, Shaon Lahiri & Anmol Narain with the assistance of Raag Bhatia (2018) Mixing and matching: using qualitative methods to improve quantitative impact evaluations (IEs) and systematic reviews (SRs) of development outcomes, *Journal of Development Effectiveness*, 10:4, 400-421, DOI: [10.1080/19439342.2018.1534875](https://doi.org/10.1080/19439342.2018.1534875)

To link to this article: <https://doi.org/10.1080/19439342.2018.1534875>



View supplementary material [↗](#)



Published online: 02 Nov 2018.



Submit your article to this journal [↗](#)



Article views: 122



View Crossmark data [↗](#)

ARTICLE



Mixing and matching: using qualitative methods to improve quantitative impact evaluations (IEs) and systematic reviews (SRs) of development outcomes

Emmanuel Jimenez^a, Hugh Waddington^b, Neeta Goel^a, Audrey Prost^c, Andrew Pullin^d, Howard White^e, Shaon Lahiri^f and Anmol Narain with the assistance of Raag Bhatia^a

^aInternational Initiative for Impact Evaluation, New Delhi; ^bInternational Initiative for Impact Evaluation, London, United Kingdom; ^cDepartment of Global Health and Development, London School of Hygiene and Tropical Medicine, London, United Kingdom; ^dSchool of Natural Sciences, Bangor University, Gwynedd, Wales, United Kingdom; ^eCampbell Collaboration, Oslo, Norway; ^fMilken Institute of Public Health, The George Washington University, Washington D.C, United States of America

ABSTRACT

Recent evaluations have begun to use qualitative data in a manner that helps improve the quality and relevance of studies through the inferences that are drawn from them, and their applicability to policy makers and programme implementers. This paper reviews this work and identifies good practices to integrate qualitative methods into quantitative impact evaluations (IEs) and systematic reviews (SRs). Using recent literature on the characteristics of such practices, we developed two tools to assess the methodological rigour and mixed methods integration of 40 IEs and 7 SRs, drawing upon previous approaches. Our findings are that successful mixed methods quantitative impact evaluations: (1) provide a clear rationale for integration of methods; (2) deploy multidisciplinary teams; (3) provide adequate documentation; and (4) acknowledge limitations to the generalisability of qualitative and quantitative findings. Successful integration tended to improve mixed methods impact evaluations by collecting better data to inform the study design and findings, which helped contextualise quantitative findings. Our main observation on the integration of mixed methods in the systematic reviews is that mixed methods systematic reviews bringing together literatures that answer different questions can go beyond the ‘sum of their parts’ to provide holistic answers about development effectiveness. The findings of this study inform several recommendations to improve the conduct and reporting of mixed methods impact evaluations and systematic reviews.

ARTICLE HISTORY

Received 17 July 2018
Accepted 4 October 2018

KEYWORDS

quantitative methods;
qualitative methods; mixed
methods; impact evaluation;
systematic review

Recent books on quantitative evaluations of development policies typically devote a chapter or two to mixed methods, or the need to complement quantitative analysis with other methods – specifically qualitative techniques (see, e.g. Gertler et al. 2016; White and Raitzer 2017). Analysts, such as Bamberger (2015), have reported how the use of mixed methods improves the overall quality of evaluations, including equity and gender-focused evaluation. However, they also say that the qualitative data incorporated into many impact evaluations tends to be limited to narrative quotes which supplement findings from the quantitative work. Similarly, in systematic reviews, qualitative findings are often brought in only during the final discussion section. Some recent

quantitative evaluations and reviews have begun to change this superficial use of qualitative data and arguably improved the quality of evaluations in terms of their relevance, the inferences that are drawn from them, and their applicability to policy-makers and programme implementers.

This paper reviews this work and identifies good practices to integrate qualitative methods into quantitative impact evaluations (IEs) and systematic reviews (SRs).¹ Using recent literature on the characteristics of such practices, we develop two tools to identify such mixed-methods IEs (MMIEs) and SRs (MMSRs), and then applied these tools to assess samples of studies from impact evaluation and systematic review databases. We focus on how well these techniques are integrated. While the similarities in ontology, epistemology and methodology in quantitatively-driven IEs and SRs ensure that our approach is consistent across these two types of studies in assessing how well the mixing of the methods is done, we also compare how the application of mixed-methods techniques may differ in each type of study.

Sections A and B of this paper discuss respectively the development of the analytical tools for MMIEs and MMSRs. Section C presents the recommendations from both parts.

1. Impact evaluations

We define quantitative impact evaluations as ‘analyses that measure the net change in outcomes for a particular group of people that can be attributed to a specific programme using the best methodology available, feasible and appropriate to the evaluation question that is being investigated and to the specific context’ (3ie [n.d.b](#)). Further, while recognising that there are many definitions of qualitative research, we use the following:

“Qualitative research is a situated activity that locates the observer in the world. It consists of a set of interpretive, material practices that make the world visible. These practices transform the world. They turn the world into a series of representations, including field notes, interviews, conversations, photographs, recordings, and memos to the self. At this level, qualitative research involves an interpretive naturalistic approach to the world. This means that qualitative researchers study things in their natural settings, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them” (Denzin and Lincoln 2018).²

Quantitative and qualitative methods must be viewed as complementary in enabling decision-makers to assess development programmes. What does it mean to have a successful integration of mixed methods techniques? There have been several interpretations of this (see, e.g. Maxwell, Chmiel, and Rogers 2015). A common approach to integrating qualitative data collection in impact evaluations involves using these data to triangulate quantitative results on effects or mechanisms described on a causal pathway, checking for mechanisms that are harder to capture through quantitative measurements, and documenting any unintended intervention consequences (Schulte-Mecklenbeck, Kühberger, and Johnson 2011).

A useful summary is provided by White (2008), who identified three main ways to combine quantitative and qualitative approaches. The first is in integrating methodologies. For example, researchers can use quantitative survey data to identify which individuals/communities could be invited to take part in a qualitative study, and use results of the survey to inform the interview guide for the qualitative work. Conversely, researchers can use qualitative data to inform the stratification of a quantitative sample, the design and acceptability of the survey questionnaire (pp. 4–5). A second type of integration involves ‘confirming/reinforcing, refuting, enriching, and explaining the findings of one approach with those of the other’ (p. 4). This includes verifying quantitative results through the qualitative approach, using qualitative work to identify issues or obtain information on variables not obtained by quantitative surveys, generating hypotheses from qualitative work to be tested through the quantitative approach, and using qualitative work to understand unanticipated results from quantitative data. A final and third type of integration involves merging the findings of the two approaches into

recommendations to improve policies and programmes. In this paper, we focus on all three types of integration.

In the rest of this section we describe how we develop the tool to assess the integration of mixed methods, apply it to a sample of MMIEs and draw implications on the path forward.

1.1. Developing a tool to assess qualitative and quantitative practices

Our primary goal was to develop a tool to assess successful integration of mixed methods in impact evaluations. However, reliable evidence is a key ingredient of successful impact evaluations (Masset et al. 2018), and assessing the rigour of individual evaluation components provides insights into the credibility of the evidence generated by mixed methods evaluations. Additionally, this allows us to explore whether studies that pass the rigour test also integrate mixed methods well. Therefore, our tool also assesses the rigour of qualitative and quantitative methods in mixed methods impact evaluations.

To assess rigour in quantitative methods, our tool draws upon the evaluation criteria from a number of sources (Miles and Huberman 1994; Bamberger, Rugh, and Mabry 2012); (Langer 2017; Hombrados and Waddington 2012). These criteria cover the domains of confirmability, credibility, transferability and utilisation. Transparency in the documentation of the analysis process, and the logical sequence leading to the interpretation of results are also key elements (Mmari and University 2006). The tool contains questions covering the clarity of the study's methods and procedures (B1); the rigour of the qualitative study in conduct (B2); reporting how researcher biases affected the study (B3); methods used to address bias (B4)³; the transferability of findings to other contexts (B5); and an assessment of whether qualitative data situated the findings within the political, institutional, cultural or social context of the study (B6)⁴

Table 1. Summary of mixed methods impact evaluation appraisal tool.

Section	Types of questions	Scoring
Preliminary information	Coder name Reference to IE report Sector focus Donor Classification of impact evaluation design (RCT/Quasi/Non-experiment)	N/A
Section A: Quantitative Aspect	A1. Clear description of the study's methods and procedures A2. Rigour of the quantitative study in conduct A3. Selection bias and confounding (if an RCT) A4. Selection bias and confounding (if a natural experiment/quasi experiment/non-experiment) A5. Post-intervention biases (motivation of participants) A6. Post-intervention biases (analysis and reporting) A7. Threats to construct and external validity A8. Reportage of statistical power	o Yes o No o Not Applicable
Section B: Qualitative Aspect	B1. Clear description of the study's methods and procedures B2. Rigour of the qualitative study in conduct B3. Reportage of assumptions, values, biases B4. Attempts to address biases B5. Transferability of results B6. (Unscored) Data situated within political, institutional, cultural or social context	o Yes o No
Section C: Integration of Mixed Methods	C1. Integration theory of change/programme or logic model explored through mixed methods C2. Integration of methods to inform study design C3. Integration of methods to inform the interpretation of findings C4. Limitations of integration C5. (Unscored) Stage(s) at which qualitative evidence is incorporated into the study. C6. (Unscored) Categorisation of the type of mixed methods study into sequential exploratory, sequential explanatory and/or convergent design.	o Yes o No

In Table 1, which summarises the MMIE appraisal tool, these are covered under Sections A and B.

We note that causal attribution is a complex activity driven by a number of factors, including but not limited to internal design validity, and external validity covering the themes of transferability and utilisation. The tool covers risks to all these themes across qualitative and quantitative domains, and scores studies on the basis of what is reported in published or grey literature.

In order to address the integration of methods, we relied on appraisal frameworks put forth by authors in the field of mixed methods evaluations such as Greene, Miles, Huberman, Bamberger, Creswell and others. Greene, Caracelli, and Graham (1989) conceptual framework of triangulation, complementarity, development, initiation, and expansion, informed the appraisal criteria put forth by Miles and Huberman (1994), and the development of methodological criteria by many researchers since. Our integration tool represents an amalgamation of these approaches. One important contribution of mixed methods studies is the use of an integrated theory of change. Therefore, Section C1 of our mixed methods impact evaluation (MMIE) tool and Section D.1 of the MMSR tool assess if sample studies used an integrated theory of change or a logic model articulating the intervention causal chain from inputs and activities to outcomes. Additionally, Section C assesses the integration of mixed methods, the study design (C2), the interpretation of findings (C3), the limitations of integration (C4), the stage at which qualitative evidence was incorporated in study design (C5) and the type of mixed methods integration (C6).

The tool was reviewed by a few subject matter experts (mixed methods research) and underwent rigorous testing by independent reviewers and feedback from experts in the field prior to finalisation. The full tool is reported in the appendix of the original CEDIL paper which is available on-line.⁵

To extract data from studies, we read through all available project documents, counting them as one study (e.g. if one study had separate reports for the qualitative and quantitative components, we counted these as one study). Where necessary, we contacted the authors to request additional details or reports on the study. For our qualitative scoring, we did not use a pre-established list of themes (in fact, not all studies even reported these themes) used within the study, as our analysis focused on the factors reported in the study rather than the primary data. Each study was scored independently by two reviewers, and discrepancies (greater than three points) in unweighted scoring were discussed and resolved. We identified the top scoring studies first by their total integration score (Section C), and then sorted the studies by overall rigour (Sections A and B together). We consider the top 12 studies (scoring 4 points or more out a maximum of 6 points on Integration) as the ones that excelled at integrating mixed methods.

1.2. *Applying the tool to MMIE studies*

Studies were identified from impact evaluation repositories (DFID, 3ie, World Bank and JPAL) spanning 20 countries, of which the majority (95%) were conducted in low and middle-income countries (L&MICs), and one high income country. Five of the studies were conducted in fragile and conflict-affected contexts (four in the Democratic Republic of Congo and one in Liberia), as defined by the World Bank.⁶ Our search was also influenced by recommendations from sectoral experts. Within the stated repositories, we selected all studies that matched our inclusion criteria, and we focused on four key sectors where mixed methods evaluations were more highly represented: education, health, financial inclusion and governance. Due to the limited number of studies matching these parameters, all studies were selected, and no sub-sampling was conducted. Where studies cited related publications or reports (such as standalone baseline reports or separate publications), these were considered to be components of the same study.

There were no geographical nor date restrictions. Only English language studies were included. We purposively selected studies to represent international development sectors with a large

Table 2. Stages at which qualitative evidence is incorporated into mixed methods IEs.

At what stage is the qualitative evidence incorporated into the study?	No. of studies	Percentage of studies
Pre-intervention	9	23
During intervention	11	28
Post-intervention	29	73
After end line analysis	8	20

Note: Some studies incorporated qualitative evidence at multiple stages in study design.

impact evaluation evidence base. The sectoral distribution was roughly equal, with a higher proportion of studies from the education sector (30%), followed by financial inclusion (25%), governance (23%) and health (23%).

In our review of studies, we found that there was no standard definition of ‘mixed methods impact evaluation’, and the integration of methods was conducted and reported in diverse ways. Therefore, our inclusion criteria comprised quantitatively-driven impact evaluations (as defined above) that reported the use of at least one qualitative method.

1.2.1. Characteristics of the studies

In terms of methodology, studies in our sample comprised randomised controlled trials (62%), quasi-experimental designs (35%) and a combination of both (3%). Only 10% studies specified a qualitative research design (e.g. ethnography). A majority of studies (58%) employed multiple techniques. Focus groups discussions were the most commonly used technique (55%), followed by key informant interviews (48%) and in-depth interviews (35%). Other techniques included life histories, contribution analysis and social mapping.

As shown in Table 2, qualitative methods were mainly used after an intervention in order to explore the implementation and its impacts (73% of studies). Studies also reported incorporating qualitative methods during an intervention (28%), before an intervention is rolled out (23%) and after endline analysis (20%).

1.2.2. Quality of the studies

Methodological Rigour and the integration of mixed methods. Most studies (81% of RCTs and 60% of quasi-experimental studies) presented balance tables to demonstrate comparability between treatment and control groups. Nearly half of the RCTs describe the process of randomisation used to allocate participants to treatment and control groups. Other sources of bias, where applicable, were addressed by fewer studies: 33% of studies addressed instances of bias due to non-adherence, 8% addressed recall bias, 18% addressed social desirability bias and 10% addressed Hawthorne effects.

While methods to account for bias were generally well described for quantitative components of the impact evaluations, fewer studies demonstrated comparable thoroughness with the qualitative components. For instance, only 20% of studies reported on the analytical framework⁷ for qualitative data; in contrast, 80% of studies described a framework for quantitative analysis. Only 38% of the studies presented information on their qualitative sampling (such as the composition of focus group participants), compared with 90% of studies describing a rationale for quantitative sampling.

Only 20% studies specifically reported any form of validity checks for their qualitative findings. In summary, studies mostly do better on quantitative rather than qualitative rigour.

Our tool also assessed (but did not score) studies on research transparency practices. While there is broad agreement among both funders and researchers on principles of transparency and open data access, the use of pre-analysis plans or providing open access to data is still not common in the international development sector. Our sample is reflective of this, as the majority of studies did not report a study protocol or pre-analysis plan (73%). It is possible that teams did have pre-analysis plans or protocols, but these were not mentioned due to word limits or space constraints. Very few studies reported publicly available study data (10% and 3% respectively for quantitative and qualitative data).

A key appraisal criterion to assess the quality of mixed methods integration was the presentation of specific inferences linked to both qualitative and quantitative data. A majority of studies (75%) provided separate data for qualitative and quantitative areas of enquiry, and brought the two together to inform study findings. However, only half of the studies provided a clear rationale for the integration of qualitative and quantitative methods. Reporting the limitations of the integration is also important, as it helps to understand what it is useful for, and what is beyond its scope to answer. However, we observe only 13% of our sample reporting limitations to the integration of methods.

About a quarter of the studies reported a divergence between qualitative and quantitative findings.⁸ Almost half of the studies (43%), report how the integration of methods influenced their policy recommendations.

We identified 12 studies that scored highly on the integration indicators, defined as scoring four or more out of six points.⁹ These 12 studies included all quantitative designs, and used a variety of qualitative techniques. Half were RCTs, and a majority (83%) used multiple qualitative techniques in their evaluation. The majority (92%) of the top scoring studies were situated in low- and middle income countries (L&MICs), with two being located in fragile and conflict-affected countries (the Democratic Republic of Congo). Nearly half of the top scoring studies were from the financial inclusion sector (42%), followed by education (33%), governance (17%) and health (8%). In the next section, we identify a few characteristics these studies.

1.3. Implications of the findings from applying the tool: characteristics of high scoring MMIEs

As numerical comparisons of the characteristics of studies, which score higher or lower on integration can only take us so far, we examine examples from the highest scoring studies in detail to provide additional information about successful integration.

1.3.1. Being rigorous in applying each method

In general, we found that studies which scored highly on quantitative and qualitative rigour, also tended to score highly on integration. For instance, 13 out of 16 studies scoring above the median integration score of 3.0 also scored above the median overall rigour (quantitative + qualitative rigour) score of 133.5. When qualitative rigour was high, it was easier to discern how well a study had integrated qualitative and quantitative components. This is unsurprising given that our sample comprised quantitatively driven impact evaluations, many of which give little credence to the qualitative component.

There were exceptions such as one evaluation, which scored highly on qualitative and quantitative rigour but did not have high scores for integration. Perhaps as a result of the complexity of evaluation components, the integration of the qualitative and quantitative lines of enquiry was less thoroughly reported as compared to the qualitative and quantitative components individually. The relationship, or level of influence between qualitative and quantitative lines of enquiry, was not described for the data-collection process (either through the process evaluation or impact evaluation), though the different components all related to the same objectives of the study. Similarly, the authors did not make efforts to explain divergent findings between their different evaluation components, nor explain the limitations of their integration (i.e. what the integrated data answer, and what they cannot answer inherent to the methods used to obtain data). As a result, this study scored less highly on integration than it did on methodological rigour. This represents a missed opportunity to engage more thoroughly with a mix of different data strands, which may have informed practical recommendations for the development of this government programme.

Approximately 23% of studies across all scoring strata had a programme/intervention theory of change, or logical framework. Doing so provided clarity on how studies integrated qualitative and quantitative components. For instance, Nisbett et al. (2016), one of the top 12 scoring studies, mapped each input and output of the causal chain to the means through which they would be

investigated, with clear indicators of the respective quantitative and qualitative methods to be used. We see this as best practice, as it demonstrates a clear link between mixed methods, and the hypothesised causal chain of a programme or intervention.

1.3.2. *Providing a clear rationale for integration*

The presentation of a clear rationale for the integration of qualitative and quantitative methods was another characteristic the exemplar studies shared. Studies that scored low on integration often included focus group discussions or key informant interviews as part of their impact evaluation, without stating a clear rationale. A few studies did not present the sampling, study design or description of the qualitative component. So, although they claimed that qualitative methods were used to corroborate quantitative findings, no analyses were presented to support this claim. In these instances, the authors often did not report qualitative results comprehensively, but only make a mention in passing that qualitative findings confirmed the quantitative findings. For example, a study evaluating the effectiveness of an education intervention on school learning outcomes did not present the rationale, research questions or objectives for qualitative enquiry, and only made a passing reference (in the Conclusion section) that interviews were conducted to assess programme uptake.

1.3.3. *Using multidisciplinary teams*

To achieve fully integrated mixed methods research, putting together a multidisciplinary team, with each member working from a discipline-specific knowledge base may not be enough. Certainly, any poorly managed team has the potential to waste resources and engender conflict. Given epistemological and ontological differences between disciplines, and between the traditional 'camps' of quantitative and qualitative researchers, a multidisciplinary team may involve interdisciplinary conflict. In this regard, it is important to convene teams, in which 'researchers work jointly using a shared conceptual framework drawing together disciplinary-specific theories, concepts, and approaches to address common problems' (Rosenfield 1992). Using a shared framework with a clear delineation of tasks that transcend individual disciplines can help bridge gaps, and lead to more robust, fully integrated mixed methods research. Among our top 12 studies, Nielsen et al. (2010) describe the composition of the evaluation team including the team's sectoral, quantitative, qualitative and mixed methods expertise.

1.3.4. *Providing adequate documentation*

Confronted by space or word limit constraints (especially in journal articles), authors often neglect to report various details on methodology, transparency or other elements. However, from a reader's perspective, this presents an incomplete picture, as it is unclear why certain study decisions were made, or what the rationale for integrating methods might be. Hence, one common element

Box 1. Illustration of a successful mixed methods impact evaluation in Zambia.

Bonilla et al. (2017) conducted a mixed methods impact evaluation of the Government of Zambia's unconditional cash transfer programme on women's decision-making and empowerment. In addition to performing well on methodological appraisal criteria for qualitative and quantitative techniques at the collection, analysis and interpretation stages, the study integrates methods in the following ways:

- The authors report how quantitative data influenced the collection of qualitative data, with a clear description of the rationale for the integration methods provided in the report.
- The presentation of qualitative and quantitative findings, and reportage of how qualitative and quantitative data led to integrated inferences or interpretations. Quantitatively, the authors find only modest increases in decision-making among women in beneficiary households, compared with the control group. However, qualitatively, women in beneficiary households actually felt more empowered as a result of the intervention, which seemed to increase overall well-being for women. By considering both strands of data, the authors find that entrenched gender norms may have been the reason for increasing empowerment in only five out of nine domains.

The authors report how mixed methods presents an opportunity for improving the measurement of empowerment, including women's decision-making indicators. In this regard, the authors' combined data led to them to question empirical measures for capturing the concept of empowerment through

among our exemplar studies is the provision of adequate documentation. This could be within a report, or through supplementary reports and/or appendices. For example (see [Box 1](#)), one of the top 12 studies, Bonilla et al. (2017), describe an evaluation of the Government of Zambia's Child Grant Program in a journal article. Despite being subject to space limitations (as defined by the journal), the study provides the information needed to understand the context and the design of both qualitative and quantitative components, as well as their integration.

1.3.5. Acknowledging limitations

Acknowledging the limits of integrating qualitative and quantitative findings facilitates a better understanding of the transferability of findings, and their implications in the policy. Limitations to the integration of mixed methods are reported in 41% of the top 12 studies. For example, A study assessing the impact of teacher training on gender norms in Northern Uganda refers to limitations in quantitative methods such as self-reported surveys, which may suffer from courtesy and social desirability bias (Chinen and Elmeski 2016).

1.4. Further implications: how has successful integration contributed to better IEs?

The value of a successful integration to evaluators lies in its ability to strengthen data collection, analysis, interpretation and policy recommendations.

1.4.1. Collecting better data

At the level of data collection, integrating qualitative and quantitative lines of enquiry lies in the use of different methods of data collection, and how they inform study design and findings. For example, a study evaluating the impact of humanitarian cash transfers used participatory techniques of data collection in conflict-affected communities to identify target beneficiaries, which, in the absence of qualitative data might not have led to nuanced findings in the Democratic Republic of Congo (American Institutes for Research 2017).

1.4.2. Validating findings through integration

By virtue of combining two or more epistemologically different approaches, studies may have to deal with divergences in findings for the same phenomena studied, and 67% of the top scorers attempted to explain contradictory observations and/or findings, if applicable. In several studies that score highly on integration, when qualitative and quantitative findings diverged, the authors discussed the reasons and implications for the differences, which informed their reported interpretation of findings (Evans et al. 2014; Chinen and Elmeski 2016; Bonilla et al. 2017; American Institutes for Research 2017).

Divergent results between qualitative and quantitative components can be critical to understanding intervention/programme effects. This is because in such cases, addressing the divergence of study findings often results in more nuanced interpretations than might be afforded by using a single method alone. They also provide authors with a sense of the limitations of certain methods for addressing evaluation questions, and how these could be mitigated through the use of complementary methods. For instance, quantitative results suggested that a cash transfer programme has no impact on local savings institutions (Haynes and Merttens 2017). However, qualitative research suggested that beneficiaries actively participated in savings groups as a direct result of the transfer. In an attempt to reconcile these findings, the study team noted that the implementing agency promoted savings groups in the intervention areas both among beneficiaries and non-beneficiaries, thereby potentially undermining programme impact.

1.4.3. Contextualising quantitative results

The use of qualitative methods can enhance the understanding of quantitative results by providing the context or background necessary to situate the findings. In cases where findings (across qualitative and quantitative methods) converge, they enhance confidence in study results. In other cases, they provide insights into the validity and limitations of quantitative findings. For instance, Langford and Panter-Brick (2013) use data obtained by ethnographic observations and in-depth interviews to understand the impact of a hand-washing intervention on women. In contrast with the quantitative findings, which demonstrated a 40% reduction in child diarrhoea, the triangulated qualitative findings highlight an important negative impact of the intervention. The qualitative findings suggest that the ultra-poor in the sample were not only unable to take-up the intervention due to structural constraints and competing priorities, but also suffered social censure from those in the sample who participated in the intervention.

Additionally, integrating qualitative methods of enquiry also helped the authors to explore unintended consequences of the intervention. For instance, in the study by the American Institutes for Research (2017), qualitative data collection indicated instances in which a cash transfer to a household reinforced male power dynamics, as some husbands used the money for alcohol or prostitutes. However, such findings were not evident in the quantitative data, potentially because respondents felt more comfortable revealing personal details in in-depth interviews rather than in quantitative surveys in the presence of other household members.

1.4.4. Contributing to forming policy recommendations

Successful integrations can inform policy recommendations by making contextually relevant policy recommendations. All of the exemplar studies report how mixed methods data influenced their policy recommendations. For example, an evaluation of a nutrition programme in Bangladesh links quantitative and qualitative findings to specific policy conclusions. Since the quantitative methods are not able to detect significant impacts of the intervention, and qualitative evidence points to specific nodes in the intervention pathway that did not lead positive outcomes, policy recommendations focus on resolving those issues (Nisbett et al. 2016).

2. Systematic reviews

Systematic reviews locate, critically appraise and synthesise all the high quality evidence from multiple contexts. Their purpose is 'to sum up the best available research on a specific question [which] is done by synthesising the results of several studies.'¹⁰ We define systematic reviews as having '(1) a well-defined question for the review, (2) an explicit search strategy, (3) clear criteria for the inclusion or exclusion of studies, (4) systematic coding and critical appraisal of included studies and (5) a systematic synthesis of study findings' (White and Waddington, 2012).

Historically, systematic reviews have used either quantitative or qualitative methods to synthesise evidence on a particular question. They usually focus on quantitative evidence from counterfactual impact evaluations to answer questions about effects. However, methods for synthesising qualitative evidence, such as meta-ethnography (Noblit and Hare 1988) and thematic synthesis including of 'barriers and facilitators' (Thomas et al. 2004) are being increasingly used by international development researchers (e.g. Munro et al. 2007; Skolidou and Oya 2018). Increasing numbers of reviews on international development topics incorporate mixed methods in 'parallel review modules' (Snilstveit 2012). These studies may combine quantitative and qualitative evidence to answer different questions relating to *programme effectiveness* along the causal pathway from programme design, implementation and targeting through to intermediate outcomes and endpoint outcomes (White 2009). In such studies, the quantitative evidence is used to provide evidence on the magnitudes of intervention effects on outcomes, while the qualitative information is typically used to provide evidence about participant or practitioner views, or to analyse alternate programme pathways or unintended

outcomes; in some cases the evidence may be combined to explore heterogeneity in findings by context, programme design and implementation (Waddington, Masset, and Jimenez 2018).

In much the same way as primary studies, qualitative evidence is often incorporated into systematic reviews of effects by drawing on a programme theory or logic model, where the qualitative evidence is used to open up the intervention ‘black box’ or assess underlying assumptions along the causal pathway, or to articulate different causal mechanisms (White 2009; Anderson et al., 2011, 2018; Kneale et al. 2018).

In this section, we describe the approach to developing a tool to appraise mixed methods systematic reviews of effects, by which we mean systematic reviews that incorporate synthesis of quantitative counterfactual evidence (often using statistical meta-analysis) alongside synthesis of qualitative evidence. The approach, analogous to that taken in the first part of the paper, aims to assess the value added by qualitative evidence synthesis in answering questions for which experimental and quasi-experimental designs may be less suited, and which other types of evidence are better placed to answer; an example given by Bamberger (2015) is the contribution of equity- or gender-focused evaluations (themselves drawing on mixed methods) in uncovering subjugated knowledge. The analysis serves to assess the extent to which mixed methods systematic reviews can provide more holistic answers than systematic reviews drawing solely on quantitative evidence from impact evaluations. We then present findings of piloting the tool on a number of systematic reviews that incorporated qualitative and quantitative evidence systematically.

2.1. Approach to assess incorporation of qualitative evidence in SRs of effects

The importance of using theory and mixed methods to develop relevant review questions, structure evidence collection, and present findings is well recognised in systematic reviewing including in international development. Snilstveit (2012) argues that mixed methods reviews are necessary to answer the important ‘what is the what?’ question for a review’s construct validity (external validity of the review to the issues at hand), by articulating at the very least the intervention design and process, which a ‘bare bones’ review drawing only on impact evaluation reports and journal articles is usually unable to do. Noyes et al. (2011) indicate reviews of effects can incorporate qualitative evidence in the following ways:

- ‘informing reviews’ in order to define the question and ensure the review includes relevant outcomes (e.g. informing the theory of change);
- ‘enhancing reviews’ by incorporating qualitative evidence contained in the impact evaluation reports;
- ‘extending reviews’ through additional searches for evidence from qualitative studies, in order to address questions about effects (e.g. exploring heterogeneity in findings); and
- ‘supplementing reviews’ in order to answer different questions through ‘a stand-alone, but complementary, qualitative review to address [different] questions.’

These roughly map onto the Greene, Caracelli, and Graham (1989) categories of sequential exploratory (informing reviews), sequential explanatory (enhancing reviews), convergent embedded (extending reviews) and convergent triangulation (supplementing reviews). We note that reviews may use multiple designs hence these categories need not be applied mutually exclusively. Furthermore, it appears common that reviews incorporating mixed methods plan to do so at the same time in the study protocol, even though the methods may be applied sequentially in study implementation itself. Therefore, we define the sequential designs as those where review modules are based on stand-alone protocols, and simultaneous designs as those where the review protocol incorporates both quantitative and qualitative review components.

In the following, we present examples of each approach for SRs.

- The purpose of *sequential explanatory design* in systematic reviews of effects is to explain quantitative results using qualitative findings. For example, the quantitative analysis is followed by qualitative evidence synthesis to explore participant views (e.g. Thomas et al. 2004), explain null findings (e.g. King, Samii, and Snilstveit 2010), or assess the applicability of findings in particular contexts (e.g. Piza et al. 2016).
- Similarly, the purpose of *sequential exploratory design* in the context of a systematic review of effects is for the qualitative findings to inform the quantitative data collection instruments and/or approach to synthesis. For example, this can take place in the development of the theory of change or determination of relevant outcomes at the systematic review protocol stage.
- Simultaneous *triangulation designs* in a systematic review of effects might triangulate the evidence through further iterations of the theory of change at the analysis phase (e.g. Carr-Hill et al. 2018), or by undertaking meta-regression analysis drawing on moderators identified from qualitative evidence synthesis (e.g. Phillips, Waddington, and White 2015).
- The purpose of simultaneous *embedded/convergent design* in the context of a systematic review of effects is to better understand a specific issue found in the quantitative meta-analysis or causal chain synthesis, using a qualitative sub-study. For example, a systematic

Table 3. Summary of mixed methods systematic reviews appraisal tool.

Section	Types of questions	Scoring
Preliminary information	Reference (SR report, protocol, summary and journal article) Date of appraisal Coder name Date of last search in review	N/A
Part 1: Critical appraisal of systematic review conduct and reporting		
Section A: Methods used to identify and include studies	Reference to a protocol Transparent reporting of review questions Specification of PICOS (separately by review question) Comprehensiveness of search Restrictions by time period/date Reporting of inclusion decisions Dependent findings	o Yes o Partially o No o Can't tell
Section B: Methods used to analyse the findings in quantitative analysis	Reporting of included studies Risk of bias (RoB) assessment Methods of analysis including effect size calculations Description of heterogeneity in findings Synthesis of findings Reporting of findings by RoB status Exploration of heterogeneity	o Yes o No o Partially o Not applicable (e.g. no included studies)
Section C: Methods used to include and analyse qualitative evidence	Searches for qualitative evidence Use of qualitative evidence Critical appraisal Reporting and analysis of findings Reporting of findings by critical appraisal	o Yes o No o Partially o Not applicable (e.g. no included studies)
Part 2: Integration of evidence and overall assessment		
Section D: Methods used to analyse the causal chain and reach conclusions	Use of programme theory Incorporation of qualitative evidence in review design Analysis of outcomes along causal chain Use of qualitative evidence in causal chain analysis Use of qualitative in other aspects Integration of qual and quant Method of reaching implications Type of mixed methods study	o Yes o No o Partially o Not applicable
Section E: Overall assessment of the reliability of the review and the incorporation of qualitative evidence	Describe the confidence in the review findings Describe how qualitative evidence is used in the review	N/A

review of effects with sub-component examining participant views (e.g. Brody et al., 2017) or implementation processes (barriers and facilitators/enabler analysis) (De Buck et al. 2017).

2.2. *Developing a tool to assess quality of MMSRs*

Drawing on the frameworks described above for mixed methods research, and existing systematic review critical appraisal tools and checklists (3ie n.d.a; Shea et al. 2017; Langer 2017), we developed a tool to categorise and critically appraise systematic reviews that incorporate quantitative and qualitative evidence to answer different questions about the effectiveness of development interventions. Table 3 summarises the tool which is available on-line in the appendix of the original CEDIL paper.¹¹

After the preliminary information section, the appraisal tool is split into five sections, the first three providing critical appraisal questions for quantitative and qualitative aspects of the review, the fourth categorising the approach to integrating the quantitative and qualitative evidence, and a final section providing a summary and overview of the full critical appraisal. The critical appraisal approach for identifying and including studies (Part 1 Section A) and methods used to analyse findings in quantitative analysis (Part 1 Section B) draw strongly on the tool used by 3ie (3ie n.d.a) itself drawing on Lewin et al. (2009) as well as Shea et al. (2017) on reporting deviations from protocol. Methods used to include and analyse findings in qualitative analysis (Section C) and methods used to analyse the causal chain and reach conclusions (Section D) are new sections. Section C draws on questions from Sections A and B. Part A includes eight questions, parts B and D include seven questions and part C includes six questions.

All of these sections have scoring (Yes, Partially, No, Can't tell, Not applicable) based on explicit decision rules. Part 2 of the tool, draws on the previous sections to describe the methods of integrating the quantitative and qualitative evidence (Section D) and provide an overall rating of the review's reliability and assessment of the use of mixed methods (Section E). Reviews are given an overall rating of overall confidence in conclusions about effects: low confidence reviews are those in which there are major methodological limitations; medium confidence reviews are those with important limitations; and high confidence reviews are those with minor limitations.

Two authors piloted the tool and revised questions and decision rules accordingly.

2.3. *Applying the tool*

2.3.1. *Sample summary*

We coded and appraised seven systematic reviews that incorporate quantitative and qualitative evidence covering agriculture, infrastructure, climate change and decentralised governance:

- King, Samii, and Snilstveit (2010) on community development approaches (governance)
- Berg and Denison (2012) on interventions against female genital mutilation/cutting (FGM/C) (public health)
- Watson et al. (2011) on energy services (infrastructure)
- Pullin et al. (2013) on protected areas (climate change)
- Waddington et al. (2014) on farmer field schools (agriculture)
- Hulland et al. (2015) on water, sanitation and hygiene (WASH) programmes (infrastructure)
- De Buck (2017) on hygiene and sanitation behaviour change (infrastructure).

We chose these reviews to illustrate the breadth of mixed methods approaches that have been applied over the years, coinciding with the advent of relatively large funding programmes for systematic reviews in international development (notably DFID and 3ie) and supported by a range of bodies (3ie, the Campbell Collaboration, the Collaboration for Environmental Evidence, and the

Evidence for Policy and Practice Information and Coordinating Centre, EPPI-centre). The reviews represent a range of review scopes, including reviews of single intervention types (e.g. farmer field schools, protected areas) and comparative reviews drawing on multiple interventions (community development, energy services, FGM/C, WASH). All reviews combined a quantitative component under a review question specifically asking about programme effects with a component drawing on qualitative evidence, whether that evidence was taken from included quantitative studies or searched for separately using explicit search and inclusion protocols. All reviews are limited to programmes implemented predominantly in, and evidence collected from, low- and middle-income countries (L&MICs). We would usually expect reviews of single interventions to be able to conduct qualitative evidence synthesis (or integrate that synthesis with quantitative evidence) at a greater depth of analysis, than comparative reviews of multiple interventions, hence conduct more comprehensive simultaneous designs. We discuss this point below. It is difficult to draw firm conclusions for a small sample.

We coded 'yes', 'partially' and 'no' responses as 1, 0.5 and 0, respectively. Overall, all included reviews were assessed as being of 'medium' or 'high confidence' in information decisions, and no reviews were assessed as being of 'low confidence'. (Figures are available on-line in the original CEDIL working paper.)

One study used a sequential explanatory design (King, Samii, and Snilstveit 2010). A second study used a simultaneous convergent design but also used a sequential exploratory design element for determining quantitative outcome categories for synthesis (Pullin et al. 2013). The remaining studies used simultaneous designs (Watson et al. 2011; Waddington et al. 2014; Hulland et al. 2015; De Buck et al. 2017). Integration of qualitative and quantitative evidence was usually based on theory of change framework and in some cases also incorporated further iterations to the theory of change or triangulation through additional quantitative analysis drawing on evidence from the qualitative synthesis (Waddington et al. 2014).

2.3.2. *Quality: methodological rigour of quantitative and qualitative analyses*

All reviews satisfied basic systematic review conduct and reporting factors including use of systematic searches of published and grey literature, explicit inclusion criteria specifying eligible populations, interventions, comparisons or contexts, outcomes and study designs (PICOS), critical appraisal of included evidence and synthesis of effects, and most satisfy internal quality assurance standards (double coding). The included reviews incorporate evidence to answer different questions along the causal chain including questions about intervention design or implementation (e.g. King, Samii, and Snilstveit 2010; Waddington et al. 2014; De Buck et al. 2017), and intermediate and endpoint outcomes (e.g. participant or practitioner views).

For general conduct and reporting (part A), quantitative synthesis (part B), qualitative synthesis (part C) and integration (part D), as we might expect, given that incorporation of mixed methods is a fairly new phenomenon in SRs, we found the average scores across reviews to decrease for each part: from 86 per cent for part A (SR standards), 78 per cent for part B (quantitative methods), 71 per cent for part C (qualitative methods) and 66 per cent for part D (integration).

On systematic review conduct and reporting, reviews tended to score best on specifying review questions and determining appropriate PICOS, and undertaking searches over an appropriate time period. Reviews scored marginally less well on reporting deviation from protocol (if any), articulating study designs for review sub-questions, avoiding bias through double-coding (at least for a random sample of included studies) and accounting for dependent findings consistently across quantitative and qualitative studies.

On the quality of quantitative evidence appraisal and synthesis, reviews tended to score highest on effect size calculation and reporting of heterogeneity. Reviews performed worse on critical appraisal (using appropriate risk of bias assessment), synthesis methods (including reporting findings by bias categories), and worst on reporting characteristics of included studies (usually due to lack of independent coding by two reviewers).

On the quality of qualitative evidence appraisal and synthesis, reviews tended to score best on methods for obtaining qualitative evidence (usually through separate systematic searches) and synthesis methods (often thematic synthesis). Reviews performed less well on reporting findings of qualitative studies accurately (due to single coding of evidence) and reporting evidence appropriately (indicating which evidence was of higher quality).

There also seems to have been an evolution in approaches to incorporating qualitative evidence over time. Early cases undertook ‘enhanced reviews’ (Noyes et al. 2011) by incorporating qualitative evidence contained in the studies eligible for the quantitative review of effects (King, Samii, and Snilstveit 2010). This approach could be a model for evidence synthesis if the impact evaluations on which reviews of effects are based typically used theory-based approaches to open up the intervention black box and present outcomes along the causal chain. However, due to early realisation that this was typically not the case (White 2009; Snilstveit 2012) mixed-methods systematic reviews have been increasingly inclusive in incorporating qualitative evidence, including by undertaking additional searches for qualitative studies linked to the included quantitative studies (Watson et al. 2011) or by conducting full searches for qualitative studies to answer specific review questions (Waddington et al. 2014).

2.3.3. How have SRs integrated qualitative analysis?

Reviews have used different methods to integrate qualitative evidence. From our small sample of studies, it seems that earlier reviews used sequential explanatory designs to incorporate qualitative evidence after the quantitative synthesis had been conducted (King, Samii, and Snilstveit 2010; Berg and Denison 2012). However, authors used different methods of synthesis. King et al. adopt a theory-based approach drawing on a logic model/theory of change and presenting evidence on implementation and processes according to that model. Berg and Denison (2012) use a realist synthesis approach to integrate the quantitative and qualitative evidence, without drawing on an explicit logic model or theory of change.

With regards the coding tool, we found that reviews tended to score most highly on incorporating qualitative evidence in the design (usually through parallel review modules) and analysis (relating to specific parts of the causal chain), and integration of quantitative and qualitative evidence (e.g. through iterations of the logic model or theory of change, or through formal statistical testing). Reviews performed marginally less well on causal chain analysis of outcomes (it is still common practice that reviews do not collect evidence on intermediate and ‘endpoint’ outcomes consistently), although the extent to which reviews should necessarily do this as best practice is debatable. For example, it is not clear whether a review of hygiene and sanitation behaviour change as a primary outcome (De Buck et al. 2017) would necessarily need to incorporate outcomes further along the causal chain as primary outcomes as well. The answer depends on the relevant question being asked in the review. Other questions in the tool relating to use of qualitative evidence in other ways, for example to assess applicability of evidence (D.5), may also need to be incorporated into previous questions (D.4) as it is doubtful whether reviews should necessarily provide this analysis. Finally, the methods in which implications for policy and practice are drawn from the quantitative and qualitative syntheses are usually limited (use of summary of findings tables, GRADE or cerQUAL approaches).

2.4. Implications for the use of qualitative methods in SRs

Reviews of qualitative evidence can answer questions about effectiveness early in the intervention causal chain, such as on intervention design and implementation or participant views. Reviews of quantitative impact evidence can answer questions about the change in net outcomes resulting from the intervention. Mixed methods reviews can go beyond the ‘sum of their parts’ to provide holistic answers about development effectiveness. In some cases, the key contribution of integration was the identification of impacts that informed the approach to analysing the quantitative

data, ensuring comprehensive consideration of evidence, even in areas where little or no quantitative evidence was found (Pullin et al. 2013). In others, the contribution has been to provide evidence on the scalability of interventions (Waddington et al. 2014). In other cases, the integration has enabled reviews to provide evidence on unintended adverse outcomes for vulnerable groups. To take some examples of reviews not included in this assessment, Brody et al. (2017) find increases in gender based violence due to economic self-help groups in the short term, and Lawry et al. (2017) find displacement of women as a result of tenure reform in Africa.

We note here that the use of mixed methods in systematic reviews should be done as appropriate to the questions being asked and the underlying evidence base being reviewed (and also cognisant of the resources available to the review team). For example, it may be possible to explain heterogeneity in impacts across studies (i.e. answer the why question) drawing solely on quantitative impact evaluations, if (and only if) these studies collect outcomes along the causal chain (an example is Welch et al. 2016). This approach, however, requires a minimum number of SRs to be identified, which we know is still not available on many topics (Cameron, Mishra, and Brown 2016). In some cases, it may be the case that sequential rather than simultaneous designs are sufficient to answer the questions being asked (a priori by helping develop the review questions and programme theory, a posteriori by helping explain findings such as impact heterogeneity). This might be the case, for example, for large comparative reviews of multiple interventions. But in other cases, such as where reviews aim to answer broader questions about a particular intervention, such as about implementation, and then link that evidence with evidence on effects, simultaneous designs are optimal.

Indeed, it is at these initial stages of the review process that formal guidance is most lacking on effective mixed methods approaches, especially convening the study team and constructing the initial conceptual framework to support the integration of qualitative and quantitative evidence. Establishing teams with appropriate qualitative and quantitative skills, preferably drawing on broad academic disciplines, is usually needed for high quality mixed methods reviews to be done efficiently. All of the reviews presented here included authors with quantitative and qualitative skills from academic disciplines including social sciences (e.g. anthropology, economics, policy science, public health and sociology) and environmental science. The reviews also drew on explicit programme theory (logic model, theory of change) to identify at what points qualitative and quantitative evidence provided the most valuable contributions to understanding the causal chain/pathways.

3. Concluding remarks and recommendations

Our paper explores the contributions of qualitative methods to quantitative impact evaluations and systematic reviews. As described in our paper, the use of mixed methods can enhance the quality of the impact evaluations by strengthening data collection, analysis and interpretation. More specifically, qualitative methods can help discern how and for whom an intervention had impacts, the mechanisms that translated inputs into certain outcomes, the trajectory of identified impacts (linear or non-linear), and also in identifying unforeseen consequences. Additionally, the complementarity between qualitative and quantitative methods can be usefully applied to systematic reviews at least to inform the questions being asked and the programme theory or logic model used, if not systematically throughout the data collection and analyses stages. Through these contributions mixed methods add value that is well beyond just the individual contributions of quantitative and qualitative findings.

We note several important limitations to this study. First, our sample was purposive and limited to 40 studies across four sectors. Hence, our results may have limited applications for other sectors or impact evaluations in general. Second, while our tool underwent intensive internal testing and refinement, including consultations with thematic experts, like any quality appraisal tool, ours could be improved with further testing. However, it builds upon a number

of existing validated tools, while adding elements (such as integration) that were not available in the original tools. Third, while we attempted to limit subjectivity by having the studies coded by two independent reviewers, it is unlikely that we eliminated it. Given the inherent subjectivity involved in quantifying the rigour and integration of MMIEs, we are confident that our credibility checks and reflexivity exercise increase the transparency of our review. Fourth, our analysis was limited to information reported in the study, which may have resulted in some inadvertent gaps. For instance, if a study did not refer to a study protocol or a pre-analysis plan, its absence would be noted in our coding sheet. However, it is possible that the study team did write a study protocol, but neglected to report it. In either case, we are confident that our recommendations will be useful to improve study reporting and conduct. Finally, our paper does not cover evaluation issues, such as complexity, which may benefit more from integrated quantitative and qualitative analysis.

Similarly, in the systematic review section too, the sample was purposive and limited to seven studies. It would be useful to develop further the tool that was piloted here by undertaking critical appraisals of more systematic reviews, preferably using double coding, and further refining critical appraisal questions. Further development of the tool should take into account the needs of decision makers.

Based on our findings, our recommendations contain elements of ‘best practice’ guidance on mixed methods integration by others (see, e.g. Creswell et al. 2011). Where we see our contribution as unique in the literature is our focus on MMIEs, and specific guidance for reporting the integration in ways that may reduce the variation in how mixed methods are conceived, implemented and reported.

First, establish a common minimum understanding of mixed methods impact evaluations (MMIEs) and mixed methods systematic reviews (MMSRs) to establish semantic consistency in the way these studies are conceived.

Based on our review, we suggest the following definitions:

A mixed methods impact evaluation (MMIE) assesses the net change in an outcome, or set of outcomes, attributed to a specific programme or intervention by comparison with a counterfactual, using complementary strands of empirical and interpretative methods to serve the evaluation question(s) being investigated. MMIEs have a clear rationale for integrating methods, and do so in at least one stage of the evaluation process in order to inform the interpretation of results.

A mixed method systematic review (MMSR) uses transparent procedures to find, evaluate and synthesise the results of quantitative and qualitative studies, in order to answer different questions. These can include questions relating to *programme effectiveness* along the causal pathway from programme design, implementation and targeting through to intermediate outcomes and end-point outcomes; to collect information on participant or practitioner views; to explore heterogeneity in findings by context, programme design and implementation; or to analyse adverse outcomes (Waddington, Masset, and Jimenez 2018).

Second, establish common minimum reporting guidelines for MMIEs and MMSRs to ensure key elements related to the integration of methods are reported. Similar to extensions to the CONSORT guidelines, we recommend reporting on a set of key elements as best practice. For MMIEs, we suggest the following integration elements:

- A clear rationale for the integration of methods as they relate to the evaluation question(s) and/or study objective(s) at each stage of the evaluation;
- A description of the composition of study teams, describing the process for establishing a common premise behind the value of mixed methods to serve the particular evaluation question(s) and/or study objective(s);

- Adequate documentation required to understand the context of research decisions behind the integration of methods, and each component within qualitative and quantitative lines of inquiry;
- Descriptions of both the quantitative and qualitative samples noting any overlaps.
- Reasonable estimates of magnitude in the presentation of results, and avoidance of imprecise language (especially when reporting qualitative results) such as ‘some participants’ or ‘most participants’.
- An acknowledgement of the ‘limitations of both qualitative and quantitative approaches in serving the evaluation question(s) and/or study objective(s);
- An acknowledgement of how researcher backgrounds, thoughts, opinions, values, and/or perspectives fit within the research process (i.e. report reflexivity);
- A presentation framework used for quantitative analysis and a plan for analysing/interpreting qualitative data, which should be reported, even if the authors are using a grounded theory approach which does not necessitate hypothesising; and
- Specifying how specific findings (and policy recommendations, if applicable) were influenced by qualitative and quantitative data; and how different qualitative and quantitative approaches were integrated for triangulation purposes.

For MMSRs, we suggest the following elements be reported:

- The rationale for integrating mixed methods as they relate to the review questions, including acknowledging the limitations of qualitative and quantitative approaches in serving the evaluation question(s) and/or study objective(s);
- A theory of change unpacking black boxes of both intervention and outcomes, as well as articulating underlying assumptions, contexts and stakeholders, in order to guide the data collection and presentation;
- Study search flow diagrams indicating the approach to sourcing quantitative evidence eligible for inclusion, and qualitative evidence eligible for inclusion. These may be separate study search flow diagrams if the qualitative searches have been conducted separately from the quantitative searches.
- Critical appraisal for each included quantitative and qualitative study, assessed using appropriate tools for quantitative and qualitative evidence.
- Separate reporting of results of quantitative and qualitative synthesis, followed by, where possible, an integrated synthesis drawing on the theory of change or other method of analysis and presentation.
- Transparent reporting of the approach used to draw conclusions (especially implications for policy and practice) from the results – e.g. summary of findings tables drawing on GRADE or cerQUAL appraisal.

Third, develop a repository of MMIEs and MMSRs that integrate qualitative and quantitative methods successfully. This repository could highlight studies where the integration of methods has led to a better appreciation of how and why an intervention works (or does not work), or where policy uptake has been high. These studies could serve as exemplars and guides for researchers who want to expand the scope of their studies beyond quantitative methods or qualitative methods alone.

Fourth, devote adequate time, monetary and human resources to designing MMIEs and MMSRs. Our findings suggest that several times, quantitative studies may include qualitative methods to explore (or corroborate) quantitative findings. However, the quantitative and qualitative researchers on a team often work and report findings independently. We suggest allocating resources to facilitate regular meetings between researchers, and have time to explore, discuss and report integrated findings and recommendations. Returning to the field to explore unexplained

differences is also a recommended best practice. Further, qualitative research can play a much more significant role in enhancing the quality (methodologically or even in terms of the findings) of the project. However, this requires careful planning of the integration of methods, ideally at different stages of the evaluation. Mapping the qualitative and quantitative components of an evaluation to the theory of change/hypothesised causal chain of a programme or intervention is a good way to conceptualise how and when both lines of enquiry should be integrated. This is especially important to avoid 'tokenistic' uses of qualitative methods, and achieve value for money when undertaking mixed methods impact evaluations.

Fifth, further develop the tools to assess MMSRs and MMIEs piloted here. We recommend that the critical appraisal of methodological rigour and integration tools developed here be used by other researchers in order to increase the reliability of the tool. Additionally, we recommend using the tool to assess qualitatively driven impact evaluations (e.g. contribution analysis) to provide a logical complement to the scope of this paper. Given that mixed methods can be time and resource intensive, we also recommend a cost-benefit analysis. Beyond further refinements to the tool, this exercise would allow us to understand the complementarity of mixed methods across a broader spectrum of impact evaluation.

Sixth, explore innovations in combining qualitative and quantitative data. A number of recent mixed methods approaches have been employed to integrate qualitative and quantitative lines of enquiry. For instance, Humphreys and Jacobs (2015) propose a unified analytical framework to aggregate across findings derived from quantitative analysis and process-based observations. In this approach, inferences from quantitative analysis may be supplemented by inferences from a smaller number of case studies by experts and adjusted to see how much they are likely to change if the expert opinions are informative or not. Glynn and Ichino (2014) suggest that when outcomes are difficult to measure, qualitative information can be converted into ordinal measures of outcomes within matched sets to reduce p-values. Additionally, QCA may be particularly useful in determining the causal constituents of complex interventions, through a fully integrated and systematic method (see also, Intervention Component Analysis by Sutcliffe et al. 2015). Similarly, realist reviews are also useful innovations that seek to understand the particularities of intervention constituents in relation to their differential effectiveness for particular populations (Pawson et al. 2005). It is worth conducting a review of these methods to understand (1) how they interpret causal attribution in MMIEs and (2) how they can add value in serving particular evaluation and review questions or contexts. In this way, they can be added usefully to a methodological tool belt for researchers and commissioners of MMIEs and MMSRs. Finally, mixed methods can play an important role in bringing together the very different approaches of data science and conventional evaluation. In our view, this is a promising area of research.

Notes

1. We recognize that the primary focus on attributable evidence is a limitation of the scope of the paper which does not mean to suggest that quantitative analysis ranks ahead of qualitative techniques.
2. This definition is consistent with that of Creswell (2014) who defines 'qualitative research' as 'a means for exploring and understanding the meaning individuals or groups ascribe to a social or human problem. The process of research involves emerging questions and procedures; collecting data in the participants' setting; analysing the data inductively; building from particulars to general themes; and making interpretations of the meanings of the data.'
3. Bias is commonly understood to be a concept drawn from the quantitative research paradigm, and incompatible with the philosophical underpinnings of qualitative enquiry (Creswell, 2014; Thorne, Stephens, and Truant 2016; Davies and Dodd 2002). Instead, qualitative researchers agree that concepts such as rigour and trustworthiness are more applicable to the subjective nature of qualitative research. Our tool incorporates these concepts based on the ideas proposed by Creswell (2014), Greene, Caracelli, and Graham (1989), Miles and Huberman (1994), Pluye et al. (2011), Langer (2017) and more.
4. Section B5 of the MMIE tool covers the description of the context and conditions under which phenomena of interest occur, and the scope and limitations of data presented to enable generalisation to other settings. The term 'thick descriptions' is typically used in ethnographies, and we erred on the side of caution by not privileging one method over the other in the scoring criteria.

5. <https://cedilprogramme.org/mixing-matching-using-qualitative-methods-quantitative-impact-evaluations/>.
6. The World Bank's harmonised list of fragile situations for 2018 is available at: <http://pubdocs.worldbank.org/en/189701503418416651/FY18FCSLIST-Final-July-2017.pdf>.
7. By analytical framework, we are referring to whether or not the study reported the themes, coding and analysis procedures.
8. For an example of such divergences, refer to Section IV, part B.
9. Integration indicators cover six domains, which include the provision of logic or programme models explored through mixed methods, the use of mixed methods to inform components of study design, and to inform the interpretation of findings, as well as limitations to the integration of methods. For more information, refer to Section C of the tool in the appendix, and part II, Section C.
10. <https://campbellcollaboration.org/research-resources/writing-a-campbell-systematic-review/systemic-review.html> (accessed 11 July 2018).
11. Please refer to footnote no. 5.

Notes on contributors

Emmanuel (Manny) Jimenez is Executive Director of the International Initiative on Impact Evaluation (3ie), a non-profit organization which provides grants for the rigorous assessment of the effectiveness of development projects and programs and supports the use of such evidence in decision-making. He came to 3ie early in 2015 after many years at the World Bank Group where he provided technical expertise and strategic leadership in a number of research and operational positions including as director of the bank's operational program in human development in its Asia regions from 2000-2012 and as director of public sector evaluations from 2012- 2014. Before joining the bank, Dr Jimenez was on the economics faculty at the University of Western Ontario in London, Canada. He received his Ph.D. from Brown University.

Hugh Waddington is Senior Evaluation Specialist in 3ie's Synthesis and Reviews Office. He has a background in research and policy, having worked previously in the Government of Rwanda, the UK National Audit Office and the World Bank, and before that with Save the Children UK and the Department for International Development. He is managing editor of the *Journal of Development Effectiveness* and co-chair of the International Development Coordinating Group (IDCG) of the Campbell Collaboration.

Neeta Goel is a Senior Evaluation Specialist in 3ie's Evaluation Office. She is responsible for the review and management of 3ie-funded research and impact evaluation grants. She has over seventeen years of experience in the international development sector. Her work includes the design, implementation and evaluation of NGO interventions focusing on disadvantaged children and communities. Prior to 3ie, Neeta worked in several national and international NGOs. In her most recent assignment, she served as the Program Director for Children International, managing programmes in ten countries across Africa, Asia and Latin America. Neeta holds a Ph.D. in Childhood Studies from Rutgers University and is a member of the American Evaluation Association.

Audrey Prost is a social anthropologist based at LSHTM, where she serves as the Head of the Doctoral College. Her research focuses on designing and evaluating community interventions to improve maternal, child and adolescent health in India. I am particularly interested in participatory interventions and community engagement methods.

Andrew Pullin is the Director of the Centre for Evidence-Based Conservation at Bangor University. He is interested in the concept of evidence synthesis and evidence-based practice in environmental management. In 2007 he co-founded the Collaboration for Environmental Evidence which promotes the conduct and dissemination of systematic reviews of evidence on environmental impacts of human actions and effectiveness of environmental management and policy interventions worldwide.

Howard White is CEO of the Campbell Collaboration. He has published widely on the effectiveness of aid and anti-poverty programmes, and evaluation and systematic review methods. He is former editor of the *Journal of Development Effectiveness* and *Journal of Development Studies*.

Shaon Lahiri is a PhD candidate in the Social and Behavioral Sciences at the George Washington University. Prior to this, he was research associate in 3ie's Evaluation Office where he worked on grants related to water, sanitation and hygiene.

Anmol Narain is a research assistant in 3ie's Evaluation Office. She provides research, planning and project management support for 3ie's thematic grant programme on promoting latrine use in rural India, and other grants related to water, sanitation and hygiene.

Acknowledgments

This paper draws on a CEDIL paper which is available on-line. We would like to acknowledge the helpful comments we received from our CEDIL and DFID reviewers as well as the anonymous reviewer for the journal without implicating them in any way for any errors in the final product.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- 3ie. **n.d.a.** *Checklist for Making Judgements about How Much Confidence to Place in a Systematic Review of Effect (Adapted Version of SURE Checklist)*. London: International Initiative for Impact Evaluation (3ie).
- 3ie. **n.d.b.** *Principles of Impact Evaluation*. New Delhi: International Initiative for Impact Evaluation (3ie).
- American Institutes for Research. **2017**. *Humanitarian Cash Transfers in the Democratic Republic of the Congo: Evidence from UNICEF's ARCC II Programme*. Washington, DC: Author.
- Anderson, L., M. Petticrew, E. Rehfuess, R., Armstrong, E., Ueffing, P., Baker, D., Francis, and P. Tugwell. **2011**. "Using Logic Models to Capture Complexity in Systematic Reviews." *Research Synthesis Methods* 2 (1): 33-42. <https://doi.org/10.1002/jrsm.32>.
- Bamberger, M. **2015**. "Innovations in the Use of Mixed Methods in Real-World Evaluation." *Journal of Development Effectiveness* 7 (3): 327–335. doi:[10.1080/19439342.2015.1068832](https://doi.org/10.1080/19439342.2015.1068832).
- Bamberger, M., J. Rugh, and L. Mabry. **2012**. *Realworld Evaluation: Working under Budget, Time, Data, and Political Constraints*. 2nd ed. California: SAGE Publications.
- Berg, R. C., and E. Denison. **2012**. "Interventions To Reduce The Prevalence Of Female Genital Mutilation/Cutting In African Countries." *Campbell Systematic Reviews* 2012: 9. doi:[10.4073/Csr.2012.9](https://doi.org/10.4073/Csr.2012.9).
- Bonilla, J., R. C. Zarzur, S. Handa, C. Nowlin, A. Peterman, H. Ring, and D. Seidenfeld. **2017**. "Cash for Women's Empowerment? A Mixed-Methods Evaluation of the Government of Zambia's Child Grant Program." *World Development* 95: 55–72. doi:[10.1016/j.worlddev.2017.02.017](https://doi.org/10.1016/j.worlddev.2017.02.017).
- Brody, C., T. De Hoop, M. Vojtkova, R. Warnock, M. Dunbar, P. Murthy, and S. Dworkin. **2017**. "Can Economic Self-Help Group Programs Improve Women's Empowerment? A Systematic Review." *Journal Of Development Effectiveness* 9 (1): 15–40. doi:[10.1080/19439342.2016.1206607](https://doi.org/10.1080/19439342.2016.1206607).
- Cameron, D., A. Mishra, and A. Brown. **2016**. "The Growth of Impact Evaluation for International Development: How Much Have We Learned?" *Journal of Development Effectiveness* 8 (1): 1–21. doi:[10.1080/19439342.2015.1034156](https://doi.org/10.1080/19439342.2015.1034156).
- Carr-Hill, R., C. Rolleston, R. Schendel, and H. Waddington. **2018**. "The Effectiveness of School-Based Decision Making in Improving Educational Outcomes: A Systematic Review." *Journal of Development Effectiveness* 10 (1): 61–94. doi:[10.1080/19439342.2018.1440250](https://doi.org/10.1080/19439342.2018.1440250).
- Chinen, M., and M. Elmeski. **2016**. *Evaluation of The Transformative Potential Of Positive Gender Socialization In Education For Peacebuilding*. Washington, D.C.: Air.
- Creswell, J. W., A. C. Klassen, V. L. Plano Clark, and K. C. Smith. **2011**. *Best Practices For Mixed Methods Research In The Health Sciences*, 2094–2103. Bethesda (Maryland): National Institutes Of Health.
- Creswell, J. W. **2014**. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. London: Sage Publications.
- Davies, D., and J. Dodd. **2002**. "Qualitative Research and the Question of Rigor." *Qualitative health research* 12 (2): 279–289. doi:[10.1177/104973230201200211](https://doi.org/10.1177/104973230201200211).
- De Buck, E., H. Van Remoortel, K. Hannes, T. Govender, S. Naidoo, B. Avau, A. Vande Veegaete, et al. **2017**. "Approaches to Promote Handwashing and Sanitation Behaviour Change in Low- and Middle-Income Countries: A Mixed Method Systematic Review." *Campbell Systematic Reviews* 2017: 7. doi:[10.4073/Csr.2017.7](https://doi.org/10.4073/Csr.2017.7).
- Denzin, N. K., and Y. S. Lincoln. **2018**. *The Sage Handbook of Qualitative Research*. 5th ed. Los Angeles: Sage.
- Evans, D., S. Hausladen, K. Kosec, and N. Reese. **2014**. *Community-Based Conditional Cash Transfers in Tanzania: Results from A Randomized Trial*. Washington, DC: World Bank Publications.
- Gertler, P. J., S. Martinez, P. Premand, L. B. Rawlings, and C. M. J. Vermeersch. **2016**. *Impact Evaluation in Practice*. Washington, DC: World Bank Group.
- Glynn, A. N., and N. Ichino. **2014**. "Using Qualitative Information to Improve Causal Inference." *American Journal of Political Science* 59 (4): 1055–1071. doi:[10.1111/ajps.12154](https://doi.org/10.1111/ajps.12154).
- Greene, J. C., V. J. Caracelli, and W. F. Graham. **1989**. "Toward A Conceptual Framework for Mixed-Method Evaluation Designs." *Educational Evaluation And Policy Analysis* 11: 255–274. doi:[10.3102/01623737011003255](https://doi.org/10.3102/01623737011003255).

- Haynes, A., and F. Merttens. 2017. *Evaluation of The Uganda Social Assistance Grants for Empowerment (SAGE) Programme (No.354)*. International Policy Centre for Inclusive Growth. Brazil: United Nations Development Programme(UNDP).
- Hombrados, J. G., and H. Waddington. 2012. *Tool to Assess Risk of Bias and Internal Validity of Social Experiments and Quasi-Experiments*. Mimeo. London: International Initiative For Impact Evaluation (3ie).
- Hulland, K., N. Martin, R. Dreibelbis, J. Debruicker Valliant, and P. Winch. 2015. "What Factors Affect Sustained Adoption of Safe Water, Hygiene and Sanitation Technologies?" In *A Systematic Review of Literature*. London: Eppi-centre, Social Science Research Unit, Ucl Institute Of Education, University College London
- Humphreys, M., and A. M. Jacobs. 2015. "Mixing Methods: A Bayesian Approach." *American Political Science Review* 109: 653–673. doi:10.1017/S0003055415000453.
- King, E., C. Samii, and B. Snilstveit. 2010. *Interventions to Promote Social Cohesion in Sub-Saharan Africa. Synthetic Review 002*. New Delhi: International Initiative For Impact Evaluation.
- Kneale, D., D. Gough, M. Bangpan, H. Waddington, and J. Thomas. 2018. *Causal Chain Analysis in Systematic Reviews of International Development Interventions*. Unpublished Manuscript. London: Centre of Excellence for Development Impact and Learning (Cedil).
- Langer, L. 2017. "Mixed-Methods Critical Appraisal Tool." PhD diss., University of Johannesburg. 1–19. Johannesburg: Africa Centre For Evidence.
- Langford, R., and C. Panter-Brick. 2013. "A Health Equity Critique of Social Marketing: Where Interventions Have Impact but Insufficient Reach." *Social Science & Medicine* 83: 133–141. doi:10.1016/j.socscimed.2013.01.036.
- Lawry, S., C. Samii, R. Hall, A. Leopold, D. Hornby, and F. Mtero. 2017. "The Impact of Land Property Rights Interventions on Investment and Agricultural Productivity in Developing Countries: A Systematic Review." *Journal of Development Effectiveness* 9 (1): 61–81. doi:10.1080/19439342.2016.1160947.
- Lewin, S., A. D. Oxman, J. N. Lavis, and A. Fretheim. 2009. "Support Tools for Evidence-Informed Health Policymaking (Stp) 8: Deciding How Much Confidence to Place in A Systematic Review." *Health Research Policy And Systems* 7 (Suppl 1): S8. doi:10.1186/1478-4505-7-S1-S4.
- Masset, E., F. Rathinam, M. Nath., and B. Wood. 2018. *Successful Impact Evaluations: Lessons from DFID and 3ie. CEDIL Inception Paper*. London: Centre of Excellence for Development Impact and Learning (CEDIL).
- Maxwell, J., M. Chmiel, and S. Rogers. 2015. "Designing Integration in Multimethod and Mixed Methods Research." In *The Oxford Handbook of Multimethod and Mixed Methods Research Enquiry*, edited by S. Hesse-Biber and R. Johnson. 1st ed. 1–70. New York: Oxford University Press.
- Miles, M. B., and A. M. Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. California: Sage Publications.
- Mmari, K., and T. J. H. University. 2006. "Using Qualitative Methods for Monitoring and Evaluation." Power point presentation presented as a part of the course, *Fundamentals of Program Evaluation* at Johns Hopkins University. Accessed 24th October 2018. <http://ocw.jhsph.edu/courses/FundamentalsProgramEvaluation/PDFs/Lecture11.pdf>
- Munro, S. A., S. A. Lewin, H. J. Smith, M. E. Engel, A. Fretheim, and J. Volmink. 2007. "Patient Adherence To Tuberculosis Treatment: A Systematic Review Of Qualitative Research." *Plos Med* 4 (7): E238. doi:10.1371/journal.pmed.0040238.
- Nielsen, N., K. Godden, P. Leguene, D. Ruegenberg, and J. Rüdiger. 2010. *Wfp Cambodia School Feeding 2000–2010: A Mixed Method Impact Evaluation*. Rome: World Food Programme Office Of Evaluation.
- Nisbett, N., R. Longhurst, I. Barnett, F. Feruglio, J. Gordon, J. Hoddinott, F. Jahan, N. Karachiwalla, S. Roy, and V. Shah. 2016. *Impact Evaluation of The Dfid Programme to Accelerate Improved Nutrition for The Extreme Poor in Bangladesh*. Brighton: Maximising the Quality of Scaling Up Nutrition Programme (MQSUN)
- Noblit, G. W., and R. D. Hare. 1988. *Meta-Ethnography: Synthesizing Qualitative Studies. Qualitative Research Methods*. Vol. 11. Newbury Park, California: Sage Publications.
- Noyes, J., J. Popay, A. Pearson, K. Hannes, and A. Booth, On Behalf of the Cochrane Qualitative Research Methods Group. 2011. "Qualitative Research and Cochrane Reviews." Chapter 20 in *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*, edited by J. P. T. Higgins and S. Green. London: The Cochrane Collaboration.
- Pawson, R., T. Greenhalgh, G. Harvey, and K. Walshe. 2005. "Realist Review – a New Method of Systematic Review Designed for Complex Policy Interventions." *Journal of Health Services Research & Policy* 10: 21–34. doi:10.1258/1355819054308530.
- Phillips, D., H. Waddington, and H. White. 2015. "Better Targeting of Farmers as a Channel for Poverty Reduction: A Systematic Review of Farmer Field Schools Targeting." *Development Studies Research* 1 (1): 113–136. doi:10.1080/21665095.2014.924841.
- Piza, C., T. Cravo, L. Taylor, L. Gonzalez, I. Musse, I. Furtado, A. C. Sierra, and S. Abdelnour. 2016. *Business Support for Small and Medium Enterprises in Low- and Middle-Income Countries: A Systematic Review, 3ie Systematic Review*. Vol. 25. London: International Initiative For Impact Evaluation.
- Pluye, Pierre, E. Robert, M. Cargo, Gillian Bartlett, A. O'cathain, Frances Griffiths, Felicity Boardman, Marie-Pierre Gagnon, and M. C. Rousseau. (2011). "Proposal: A mixed methods appraisal tool for systematic mixed studies reviews." 1–8. Montréal: McGill University.
- Pullin, A. S., M. Bangpan, S. Dalrymple, K. Dickson, N. R. Haddaway, J. R. Healey, H. Hauari, et al. 2013. "Human Well-being Impacts of Terrestrial Protected Areas." *Environmental Evidence* 2 (19): 1–41. doi:10.1186/2047-2382-2-19.

- Rosenfield, P. L. 1992. "The Potential of Transdisciplinary Research for Sustaining and Extending Linkages between the Health and Social Sciences." *Social Science & Medicine* 35: 1343–1357. doi:[10.1016/0277-9536\(92\)90038-R](https://doi.org/10.1016/0277-9536(92)90038-R).
- Schulte-Mecklenbeck, M., A. Kühberger, and J. G. Johnson, eds. 2011. *A Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide*. Psychology Press.
- Shea, B., B. C. Reeves, G. Wells, M. Thuku, C. Hamel, J. Moran, D. Moher, et al. 2017. "Amstar 2: A Critical Appraisal Tool for Systematic Reviews that Include Randomised or Non-Randomised Studies of Healthcare Interventions, or Both." *British Medical Journal* 358: J4008. doi:[10.1136/bmj.j4008](https://doi.org/10.1136/bmj.j4008).
- Skalidou, D., and C. Oya. 2018. "The Challenges of Screening and Synthesising Qualitative Research in a Mixed-Methods Systematic Review. The Case of the Impact of Agricultural Certification Schemes." *Journal of Development Effectiveness* 10 (1): 39–60. doi:[10.1080/19439342.2018.1438495](https://doi.org/10.1080/19439342.2018.1438495).
- Snilstveit, B. 2012. "Systematic Reviews: From 'Bare Bones' Reviews to Policy Relevance." *Journal of Development Effectiveness* 4 (3): 388–408. doi:[10.1080/19439342.2012.709875](https://doi.org/10.1080/19439342.2012.709875).
- Sutcliffe, K., J. Thomas, G. Stokes, K. Hinds, and M. Bangpan. 2015. "Intervention Component Analysis (ICA): A Pragmatic Approach for Identifying the Critical Features of Complex Interventions." *Systematic Reviews* 4: 140. doi:[10.1186/s13643-015-0126-z](https://doi.org/10.1186/s13643-015-0126-z).
- Thomas, J., A. Harden, A. Oakley, S. Oliver, K. Sutcliffe, R. Rees, G. Brunton, and J. Kavanagh. 2004. "Integrating Qualitative Research with Trials in Systematic Reviews." *British Medical Journal* 28: 1010. doi:[10.1136/Bmj.328.7446.1010](https://doi.org/10.1136/Bmj.328.7446.1010).
- Thorne, S., J. Stephens, and T. Truant. 2016. "Building Qualitative Study Design Using Nursing's Disciplinary Epistemology." *Journal of advanced nursing* 72 (2): 451–460. doi:[10.1111/jan.2016.72.issue-2](https://doi.org/10.1111/jan.2016.72.issue-2).
- Waddington, H., B. Snilstveit, J. G. Hombrados, M. Vojtkova, D. Phillips, P. Davies, and H. White. 2014. "Farmer Field Schools for Improving Farming Practices and Farmer Outcomes in Low-And Middle-Income Countries: A Systematic Review." *Campbell Systematic Reviews*, 1–336.
- Waddington, H., E. Masset, and E. Jimenez. 2018. "What Have We Learned From 10 Years of Systematic Reviews in International Development?" *Journal Of Development Effectiveness* 10 (1): 1–16. doi:[10.1080/19439342.2018.1441166](https://doi.org/10.1080/19439342.2018.1441166).
- Watson, J., R. Byrne, J. Opazo, F. Tsang, M. Morgan-Jones, and S. Diepeveen. 2011. "What Are the Major Barriers to Increased Use of Modern Energy Services Among the World's Poorest People and are Interventions to Overcome These Effective?" *Cee Protocol 11-004. Collaboration For Environmental Evidence*. www.Environmentalevidence.Org/Sr11004.Html
- Welch, V. A., E. Ghogomu, A. Hossain, S. Awasthi, Z. A. Bhutta, C. Cumberbatch, R. Fletcher, et al. 2016. "Deworming and Adjuvant Interventions for Improving the Developmental Health and Well-Being of Children in Low- and Middle-Income Countries: A Systematic Review and Network Meta-Analysis." *Campbell Systematic Reviews* 2016: 7. doi:[10.4073/csr.2016.7](https://doi.org/10.4073/csr.2016.7).
- White, H. 2008. "Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation." *IDS Bulletin* 39: 98–109. doi:[10.1111/j.1759-5436.2008.tb00436.x](https://doi.org/10.1111/j.1759-5436.2008.tb00436.x).
- White, H. 2009. "Theory-Based Impact Evaluation: Principles and Practice." *Journal of Development Effectiveness* 1 (3): 271–284. doi:[10.1080/19439340903114628](https://doi.org/10.1080/19439340903114628).
- White, H. 2018. "Theory Based Systematic Reviews." *Journal Development Effectiveness* 10: 1. doi:[10.1080/19439342.2018.1439078](https://doi.org/10.1080/19439342.2018.1439078).
- White, H., and D. A. Raitzer. 2017. *Impact Evaluation of Development Interventions: A Practical Guide*. Manila: Asian Development Bank. doi:[10.22617/Tcs179188-2](https://doi.org/10.22617/Tcs179188-2).
- White, H., and H. Waddington. 2012. "Why Do We Care about Evidence Synthesis? An Introduction to The Special Issue on Systematic Reviews." *Journal of Development Effectiveness* 4 (3): 351–358. doi: [10.1080/19439342.2012.711343](https://doi.org/10.1080/19439342.2012.711343).